Journal of Forestry Research (2008) 19(3):231–234
DOI 10.1007/s11676-008-0037-0

RESEARCH PAPER

Re-engineering of the forest stand database: case study of Bilahe Forestry Bureau, Inner Mongolia of China

CHEN Wen-bo[1*], ZHENG Jiao[2]

[1] *Research Center of Landscape and Environment Ecology, Jiangxi Agricultural University, Nan Chang 330045, P.R. China*
[2] *College of Computer and Information Engineering, Jiangxi Agricultural University, Nan Chang 330045, P.R. China*

Abstract: The forest stand database of Bilahe Forestry Bureau, Inner Mongolia of China was taken as an example to demonstrate the whole process of building a temporal geodatabase by means of reengineering. The process was composed of establishing a conceptual data model from the initial database, constructing a logical database by means of mapping, and building a temporal geodatabase with the help of Computer-Aided Software Engineering (CASE) tool and Unified Markup Language (UML). The results showed that as the reengineered forest stand geodatabase was dynamic, it could easily store the historical data and answer time related questions by Structured Query Language (SQL), meanwhile, it maintains the integrity of database and eliminates the redundancy.

Keywords: re-engineering; spatial-temporal database; CASE tool; forest stand

Introduction

Organizations are turning to system reengineering as a means of upgrading their existing information systems in situations where it appears to be a less expensive alternative to system replacement. Re-engineering can be defined as the process of discovering how a system works. It requires identifying and understanding all components of an existing system and the relationships between them (Alhajj 2003). Furthermore, reengineering is necessary to semantically enrich and document a database in order to avoid throwing away huge amounts of data stored in existing legacy database if the owner of an existing database wants to maintain or adjust the database design (Alhajj 2003; de Guzman et al. 2006).

The re-engineering process derives the conceptual schema from the existing database and the objective is to extract and know as much necessary information as possible about the conceptual model that led to the legacy database being re-engineered, namely, from legacy database to it higher conceptual data model. There is no universal agreement on how to do the database re-engineering. Although methods vary from person to person, they are all under the guidelines of reducing redundancy, eliminating null values and maintaining the integrity of database (Cohen and Feldman 2003).

More and more organizations prefer to real-world spatial database with accountability and traceability. These requirements lead to the replacement of the usual ‘update-in-place’ policy by an ‘append-only’ policy that retains all previous records in the database, namely, transaction-time policy (Sotnykova et al. 2005; Zhao et al. 2004). Various temporal data models and corresponding temporal data dependencies and temporal normal forms have been proposed. Although these temporal models are very diverse, they can be grouped into two types: tuple time-stamping (for example Dey et al. 1996) and attribute time-stamping (for example Liao et al. 1999).

The importance of spatial temporal database is not only recognized by database engineers, but also by database users and practitioners.

Problem identification and objectives of the study

Bilahe Forestry Bureau, Inner Mongolia, China is located in Daxing’anling mountainous area of northeast of China. It covers an area of 47 000 ha, most of which is densely forested by three tree species, i.e., *Larix gmelini, Quercus mongolica* and *Betula platyphylla.* Three large scale forest inventories were conducted in 1980s, 1990s and 2000s, respectively, since its foundation. Along with yearly logging plan, the Bureau carries out a complementary inventory every year. The management units are classified into three levels, e.g. forest stand, forest compartment and forest farm. Each forest farm can have one to many forest compartments that are composed of one to many forest stands. The data of forest inventory are detailed to each forest stand, and then summed up to higher management units.

Bilahe Forestry Bureau was mainly logging-oriented and the

Foundation project: China Scholarship (2004836052); ‘Voyage Project’ of Jiangxi Province(2007)
Received: 2008-01-14; Accepted: 2008-03-18

Biography: CHEN Wen-bo (1974-), male, Ph.D., associate professor of College of Land Resources and Environment, Jiangxi Agricultural University, Nan Chang 330045, P.R. China. Email: cwb1974@126.com
Responsible editor: Hu Yanbo

natural forest landscape had been dramatically damaged due to the gradually stronger human intervention. In the early of 1999, the "natural forest restoration" project after the unhistorical deluge of 1998 in China came into being, aiming to protect the natural forest of Daxing'anling forest region. Since then, Bilahe Forestry Bureau shifted its function from logging-oriented to ecology-oriented. Temporal information plays a crucial role in such a case. The current inventory data are loosely managed by Microsoft Excel. They contain a group of 12 spreadsheets of each year, each of which distinguishes itself a major topic relevant to forest stand (e.g. terrain, soil, regeneration, disease). The spreadsheets are linked by the unique identification of forest stand. Temporal information is maintained by different spreadsheet of the same topic. Although there are forest stand map and forest resources distribution map, they are in hard copy instead of spatial data format. Many problems of the current database have been identified, some of which are 1) lack of data integrity, 2) lack of a rational database schema and 3) difficult to retrieve temporal and spatial information.

The decision-makers of Bilahe Forest Bureau are long-timely embarrassed by lack of a forest stand database which can provide temporal and spatial information required for better forest resources management. The main objective of this study is to reengineer the current database to facilitate the spatial and temporal data retrieval. The sub-objectives are 1) to establish a rational entity-relation model, 2) to reengineer the current database schema in accordance with the protocols of relational data model and 3) to build a temporal geodatabase in ArcGIS.

Process of re-engineering

The process of re-engineering is composed of establishing a conceptual data model from the initial database, constructing a logical database by means of mapping and building a temporal geodatabase with the help of CASE tool and UML.

Conceptual data model building by Entity-relation Diagram (ERD)

Entity relation model is one of the most commonly used tools to simplify the relational real world, describing data as entities, associations and attributes (Chen and Lu 1997). There is no absolute rule of how to define an entity. The most obvious characteristics of the model are comparatively independent existence. It is also better to take the stability into consideration when defining an entity. The entities identified in this study were forest farm, compartment and forest stand because these forest management units are comparatively independent of each other and comparatively stable over time. Attributes in ER model can be simple or composite, single-valued or multi-valued and stored or derived. In this study, attributes were generally borrowed from the original database after the process of normalization to eliminate the derived attributes, and to separate the repeating group (multi-value attribute), partial functional dependencies and the transitive functional dependencies. Furthermore, the cardinality of associations were also determined based on the user's view on database, including order of the relationship (1:1, 1:M or M:N) and the optimality (may or must).

For the sake of simplicity, tuple time-sampling strategy was adopted to represent time. For the entities, we stamp them when they come into being (Attribute "from"), which is a point event. For the associations, we identify "belong" relationship as an interval time envent (for example, stand A belongs to compartment B from year X until year Y). The change of associations was regarded as point time event (for example, stand A's ID changed from 1 001 to 10 001 in year 2000). We also employed the idea of temporal dependency, which was derived from the stability of entity (Liu and Song 2000) and expressed by weak entity. Weak entities have no key attributes of their own but a partial key, which is an or a set of attributes that can uniquely identify weak entities together with its identifier's primary key. Weak entities are identified in this case mainly because entities with different stability over time should be separated for sake of maintaining temporal dependency. For weak entities, we assumed that inventory was a point time event although inventory took time. Each weak entity has a set of attributes of the same stability. Finally, 10 entities, 5 associations and relevant attributes in this study were identified (see Fig. 1, only time attributes are present in order for simplicity).

Logical database building by mapping

Mapping is the process of translating ER conceptual data model to relational data model (RM). The major processes of mapping entity types are 1) turning each entity type into a relation, choosing an appropriate key as the primary key of the relation and 2) for any weak entity type, also creating a relation. The only difference in treatment is that in the relation we should also include the attributes that form the key of owner entity type. The key of the relation is formed together with the partial key. The mapping of association types depends largely on their cardinality. Basically, 1:1 and 1:N association types can be accommodated by adding attributes to the relation of one of the participating entity types. For an N:M association type R, clearly posting a foreign key in either of the two participating entity types is not valid and a new relation must be added to the database schema.

Geodatabase with CASE tool and UML

With the release of ArcGIS, ESRI introduced a new object-relational model called geodatabase. This new model is implemented as extension to the standard relational model by integrating it with object-oriented concepts in a manner that allows geographic objects to be modelled with their behaviors. Geodatabase supports many object-oriented concepts such as inheritance, encapsulation, polymorphism etc, providing database developers the ability to build more complex behaviors into objects. ArcGIS automatically provides the object-relational mapping and manages the integrity of the data within database. There are three general strategies to create geodatabase, namely, migrating existing databases to geodatabase, using tools in ArcCatalog or ArcToolbox to create geodatabase and using UML (Unified Markup

Language) and the CASE (Computer-Aided Software Engineering) tools to build geodatabase. Applying CASE tool for geodatabase design has become increasing popular recently. The general strategy for using UML and CASE tool to design and create geodatabase involves 3 steps. Firstly, a geodatabase will be designed in UML model (see Fig. 2). Then it is exported to XML (Metadata Interchange file) or Microsoft Repository. Finally, Schema Wizard in ArcCatalog (CASE tool) will be applied to create data schema (see Fig. 3).

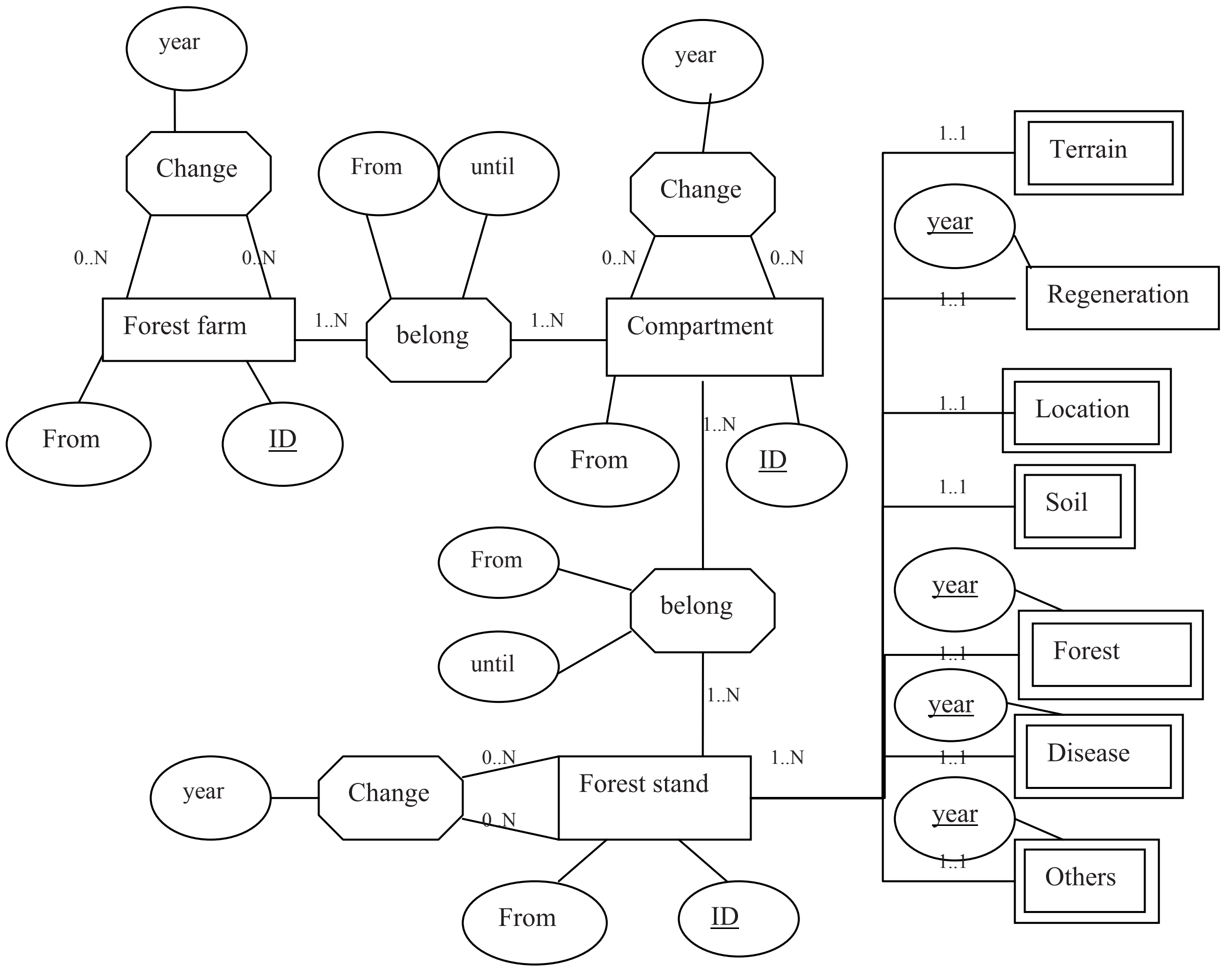


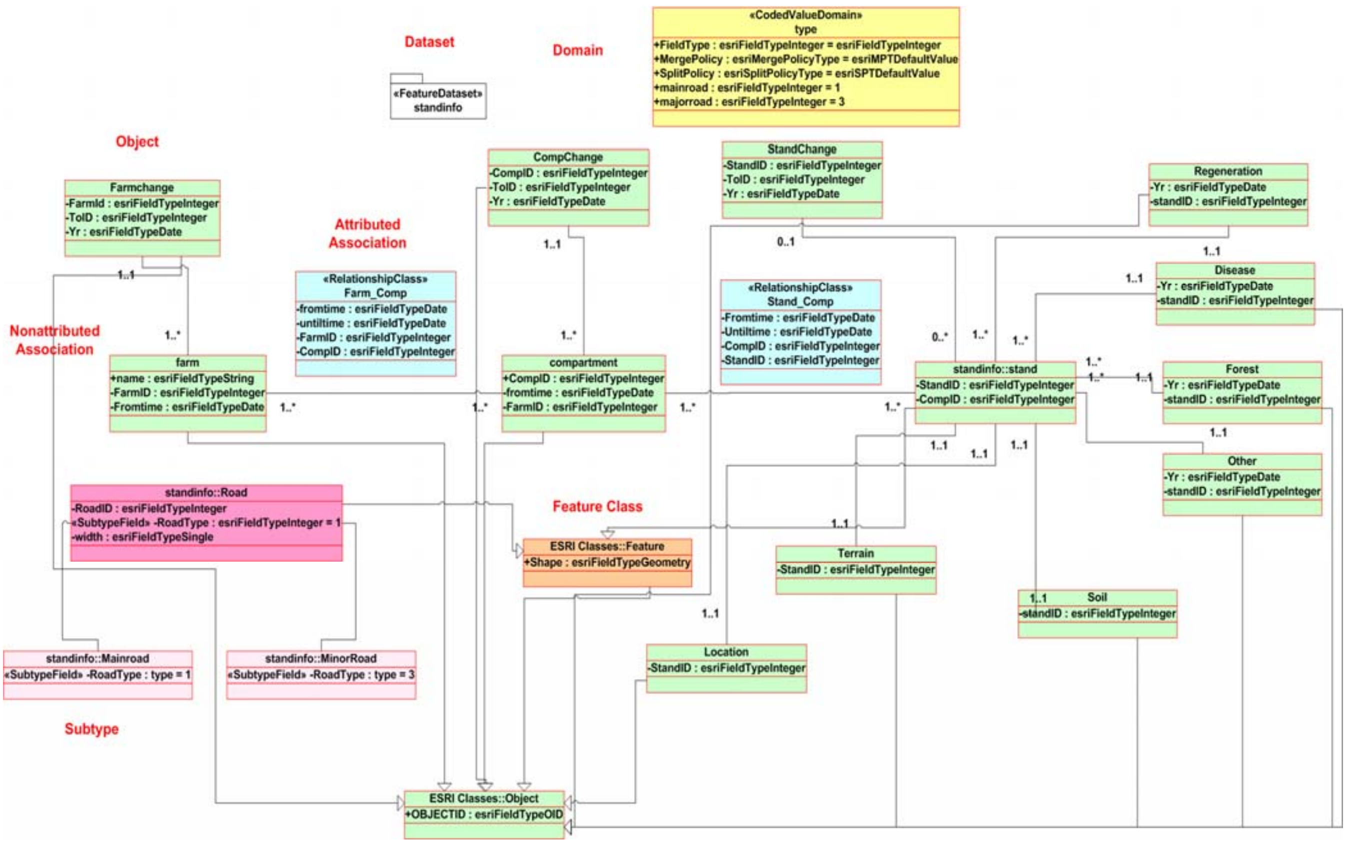
Fig. 1 Entity-relation diagram of database

Conclusions and discussions

As the Forestry Bureau of Bilahe shifted its mission from logging-oriented to ecology-oriented, it is more desirable to keep all historical inventory data in an integrated temporal database to facilitate data retrieval. The methodology of temporal database (TDB) and technology of information system re-engineering were applied to fulfil this task. During the process of re-engineering, a conceptual relational data model was re-built and expressed by Entity-relation Diagram. The tuple time-sampling strategy was adopted to represent time in database and weak entities were defined to maintain temporal dependency. Furthermore, the data dictionary (ER diagram, database schema, etc.) produced in the process of reengineering makes operation and maintenance of database easier.

In order to express history inventory information spatially, a relational object-oriented geodatabase was also built using CASE tool and UML. The most important characteristics of the reengineered database are that it is a "dynamic" database, which comprises time dimension and can easily express in database the change event of entities, associations and attributes. They can keep all the historical data in an integrated database so that it is easier to maintain database's integrity and to answer time related questions by SQL.

The reengineered database is a relational database and can be applied by any database management system (DBMS) which follows the protocol of relational database. It should be realized that a well designed database is the first step to well operated information system. In order to manage forest resources properly, a well tailored information system is also required to develop. How to develop a forest stand information system is out of the scope of this study, but it is equally crucial for forest resources management in Bilahe Forestry Bureau.



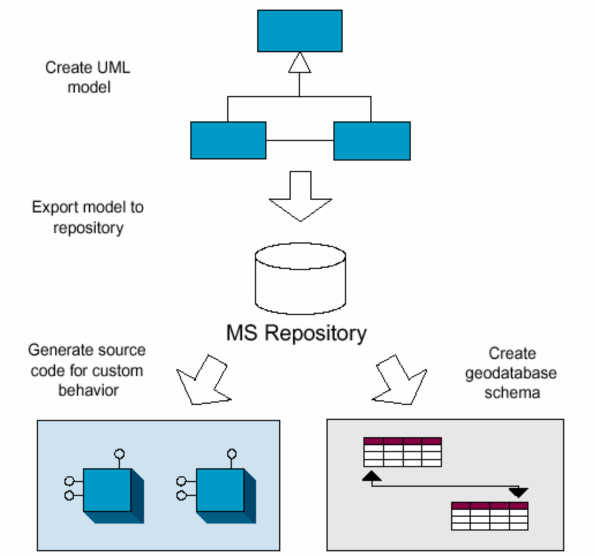


Fig. 3 The steps of creating geodatabase with the help of UML and CASE tool

Reference

Alhajj R. 2003. Extracting the extended entity-relationship model from a legacy relational database. *Information System*, **28**: 597–618.

Chen KY, Lu SS. 1997. A Petri-net and entity-relationship diagram based object-oriented design method for manufacturing systems control. *International Journal of Computer Integrated Manufacturing*, **10**(1-4): 17–28.

Cohen Y, Feldman YA. 2003. Automatic high-quality reengineering of database programs by abstraction, transformation and reimplementation. *ACM Transactions on Software Engineering and Methodology*, **12**(3): 285–316.

de Guzman IGR, Polo M, Piattini M. 2006. A methodology for database reengineering to Web Services, Model Driven Architecture - Foundations and Applications. In: Lecture Notes in Computer Science, **4066**: 226–240.

Dey D, Barron TM, Storey VC. 1996. A complete temporal relational algebra. *The VLDB journal*, **5**(3): 167–180.

Liao SY, Wang HQ, Liu WY. 1999. Stability constraints and stability normal forms for temporal relational databases. *Information and Software Technology*, 41: 131–138.

Liu WY, Song N. 2000. A logical design method for temporal databases based on the stability constraint and the generalization semantics. *Information Sciences*, **124**: 153–171.

Sotnykova A, Vangenot C, Cullot N, Bennacer N, Aufaure MA. 2005. Semantic mappings in description logics for spatio-temporal database schema integration. *Journal on Data Semantics III (Lecture Notes in Computer Science)*, **3534**: 143–167.

Zhou Xiaoguang, Chen Jun, Jiang Jie, Zhu Jianjun, Li Zhilin. 2004. Event-based incremental updating of spatio-temporal database. *Journal of Central South University of Technology*, **11**(2): 192–198.